

# SPECIFICATION

Electronic Version 1.2.8

Stylesheet Version 1.0

## **A DRAM-BASED SEPARATE I/O MEMORY SOLUTION FOR COMMUNICATION APPLICATIONS**

### Background of the Invention

[0001] Field of the Invention

[0002] The present invention generally relates to integrated circuits and, more particularly to input/output (I/O) architecture, and read/write systems for high bandwidth semiconductor memories.

[0003] Description of the Related Art

[0004] Increasing the speed of operation and bandwidth continue to be one of the incentives behind evolutionary changes in the design, development, and implementation of memory circuits. For example, for dynamic random access memories (DRAMs), the extended data output (EDO) architecture was developed to enhance memory bandwidth. With synchronous DRAMs (SDRAMs), the ability to channel the data as well as data pre-fetch schemes have helped increase the speed, throughput, and bandwidth of the memory. For example, the quad data rate (QDR) SRAM uses a four-bit pre-fetch technique to quadruple the bandwidth of the memory circuit. However, this increase in bandwidth is often accompanied by a significant increase in the overall size and cost of the memory device.

[0005]

As mentioned, one of the drawbacks of conventional QDR SRAMs is the corresponding increase in the overall size, and specifically, the number of I/O interconnect lines required to process the quad bits during read and write operations.

For example, in a QDR SRAM having a by N organization (or xN, such as x16 or x32), a 4-bit pre-fetch results in 4N bits of data being output from the memory array in read mode. Usually, the total columns in an array are divided into two sets of even and odd columns, each delivering N bits of data. The 2N columns connect to corresponding 2N sense amplifiers, with the 2N sense amplifiers driving 2N global I/O lines. Thus, a 2:1 parallel to serial conversion at the output takes place prior to the data being applied to the N data output (DQ) registers. This process is reversed for the write mode of operation, wherein a serial-to-parallel conversion turns a serial bit into quad bits for QDR.

[0006] The ability to achieve the cycle time associated with SRAM technology by using a DRAM has not yet been achieved. Those skilled in the art acknowledge that one can achieve a 4x density improvement (bit density) using DRAM over SRAM. For example, a 16 Mb SRAM approximates to a 128 Mb DRAM. However, a DRAM is much slower than its SRAM counterpart. Thus, it would be very beneficial to utilize DRAM technology due to its increased memory capacity, but to increase the DRAM speed in order to compete with SRAM processing speed. Thus, the ability to achieve the cycle time associated with SRAM technology by using a DRAM is desirable.

[0007] Separate I/O QDR SRAMs are the preferred solution in the networking/communications SRAM-memory space. Separate I/O allows read and write buses to be loaded simultaneously as memory operations constantly swap between read and write cycles. Figure 1 shows a conventional SRAM QDR timing diagram with an 8ns cycle, and a burst of 4. Moreover, conventional DRAM banking solutions use any number of banks to improve cycle time and data rate by:

[0008] DRAM Cycle

[0009] # of Banks

[0010]

For example, a 16ns DRAM with two banks can provide an 8ns cycle part. Figure 2 illustrates a conventional two-bank DRAM QDR timing diagram with an 8ns cycle, and a burst of 4. As shown, banking protocols exist which partition the memory array into several different arrays, which can then be independently addressable. That is, the partition of the memory array occurs in the bank. Conventionally, as illustrated in

Figure 1, in an 8ns timing protocol, a first bank (Bank 1) is accessed. After an 8ns cycle lapses, only then can the same bank (Bank 1) be accessed.

[0011] Thus, in a single bank solution as illustrated in Figures 1-2, the same bank may be accessed only every 8ns (the length of the entire cycle time), and during the cycle, only other banks may be accessed, not the same bank. By utilizing a multi-bank approach, for example, four banks, bandwidth maximization is achieved; as such data throughput is maximized. However, the disadvantage of such multi-banking approaches is it breaks away from the desired randomness, which is sought in banking address operations. Therefore, there is a need to access the same bank back-to-back using a multi-bank approach.

[0012] The advantage of DRAM as a replacement of SRAM memory space is becoming very popular as merged logic/DRAM processes and novel circuit techniques have enabled DRAM cycle times to shrink considerably. The need for separate I/O SRAMs with a QDR protocol has emerged as one of the standards for communication applications. Thus, there remains a need for a DRAM-based solution for the Quad Data Rate timing protocol, which allows access to the same bank during back-to-back operations.

## Summary of the Invention

[0013] The invention provides a DRAM-based separate I/O memory solution for communication applications. The present invention provides a DRAM-based solution to the standard QDR communications-SRAM solution. The present invention provides data rate and cycle times comparable to conventional QDR SRAMs. The present invention achieves at least four times the density and price/bit of conventional QDR SRAM implementations. The present invention provides a novel banking system allowing back-to-back operations to be performed on a given bank.

[0014] There is provided, according to one aspect of the invention a method of performing back-to-back read and write memory operations to the same DRAM bank, wherein the method comprises articulating between reading data on a first bank during successive first bank read cycles and writing data to a second bank during successive second bank write cycles, cycling between reading data on the second

bank during successive second bank read cycles and writing data to the first bank during successive first bank write cycles, and performing a refresh cycle on the first and second bank, wherein the first bank write cycles lag the first bank read cycles, and wherein the second bank write cycles lag the second bank read cycles.

[0015] The step of articulating comprises reading data on the first bank during the first bank read cycle, writing data to the second bank during the second bank write cycle, reading additional data on the first bank during the first bank read cycle, and writing additional data to the second bank during the second bank write cycle. The step of cycling comprises reading data on the second bank during the second bank read cycle, writing data to the first bank during the first bank write cycle, reading additional data on the second bank during the second bank read cycle, and writing additional data to the first bank during the first bank write cycle.

[0016] A refresh cycle is performed on the first bank after the step of writing additional data to the second bank. Additionally, a refresh cycle is performed on the second bank after the step of writing additional data to the first bank. The read/write memory operations constantly swap between the read and write cycles. Moreover, the read/write memory operations constantly swap between the first and second bank. Also, the read/write memory operations occur in a frequency of  $2n$  times within the same bank. Furthermore, the lag comprises at least two write cycles.

[0017] The present invention also provides a method of performing back-to-back read/write memory operations to a same DRAM bank, wherein the method comprises reading data on a first bank during an  $n$ th read cycle, writing data to a second bank during a  $n$ th write cycle, reading data on the first bank during a  $(n + 1)$  read cycle, writing data to the second bank during a  $(n + 1)$  write cycle, refreshing data on the first bank, reading data on the second bank during a  $(n + 2)$  read cycle, writing data to the first bank during a  $(n + 2)$  write cycle, reading data on the second bank during a  $(n + 3)$  read cycle, writing data to the first bank during a  $(n + 3)$  write cycle, and refreshing data on the second bank, wherein the read/write memory operations occur in a frequency of  $2n$  times within the same bank.

[0018] Additionally, according to the present invention, a method of accessing banks of dynamic random access memory (DRAM) memory is provided, wherein the method

comprises applying a first group of access requests to a first DRAM memory bank during a first time period, wherein at least one of the access requests is buffered; and applying a second group of access requests to a second DRAM memory bank during a second time period, wherein access requests that are buffered during the first time period are executed during the second time period. The first and second group of access requests comprise read/write memory operations constantly swapping between read and write cycles, wherein the read/write memory operations occur in a frequency of  $2n$  times within each of the first DRAM memory bank and the second DRAM memory bank, respectively.

[0019] The present invention further comprises a multi-bank memory circuit operable to allow back-to-back read/write memory operations to be performed on a same DRAM bank, wherein the circuit comprises a plurality of independently addressable DRAM memory banks, a refresh counter connected to the DRAM memory banks, a read register connected to the DRAM memory banks, a write register connected to the DRAM memory banks, and an address comparator connected to the read and write registers, wherein the refresh counter is operable to send an address to each DRAM memory bank, wherein the read register is operable to send a read address to each DRAM memory bank, wherein the write register is operable to send a write address to each DRAM memory bank, and wherein the address comparator is operable to compare the read and write addresses.

[0020] The circuit further comprises a partition array located in each DRAM memory bank, and an input clock operatively connected to the read and write registers, wherein the input clock is operable to capture the data from the read and write registers. Moreover, the address comparator monitors the write registers to determine if the write address has been stored in the partition array. Also, the refresh counter is operable to track all addresses to be refreshed. Moreover, the read/write memory operations constantly swap between a read and write cycle. The plurality of DRAM memory banks comprises a first and second DRAM memory bank, wherein the read/write memory operations constantly swap between the first and second DRAM memory banks. Furthermore, the read/write memory operations occur in a frequency of  $2n$  times within the same bank.

[0021] There are several benefits of the present invention. For example, the present invention provides a novel DRAM-based separate I/O memory solution for communication applications. Also, the present invention provides a novel DRAM-based solution to the standard QDR communications-SRAM solution. Additionally, the present invention provides data rate and cycle times comparable to conventional QDR SRAMs. Furthermore, the present invention achieves at least four times the density and price/bit of conventional QDR SRAM implementations. Another benefit of the present invention is that it provides a novel banking system whereby back-to-back operations are performed on a given bank.

### Brief Description of the Drawings

[0022] The invention will be better understood from the following detailed description of a preferred embodiment(s) of the invention with reference to the drawings, in which:

[0023] Figure 1 is a schematic illustration of a conventional QDR SRAM timing diagram;

[0024] Figure 2 is a schematic illustration of a conventional two-bank QDR DRAM timing diagram;

[0025] Figure 3 is a schematic illustration of a QDR timing diagram according to the present invention;

[0026] Figure 4 is a schematic illustration of a QDR timing diagram according to the present invention;

[0027] Figure 5 is a schematic block diagram of a multi-bank memory system according to the present invention;

[0028] Figure 6 is a flow diagram illustrating a preferred method of the invention;

[0029] Figure 7 is a flow diagram illustrating a preferred method of the invention;

[0030] Figure 8 is a flow diagram illustrating a preferred method of the invention;

[0031] Figure 9 is a flow diagram illustrating a preferred method of the invention;

[0032] Figure 10 is a flow diagram illustrating a preferred method of the invention;

[0033] Figure 11 is a system diagram according to the present invention.

## Detailed Description of Preferred Embodiment of the Invention

[0034] As mentioned, there remains a need for a DRAM-based solution for the Quad Data Rate timing protocol. Referring now to the drawings, and more particularly to Figures 3 through 11, there are shown preferred embodiments of the methods and structures according to the present invention.

[0035] A DRAM solution is most useful as a replacement to QDR-SRAM if back-to-back operations can be performed on the same bank. The present invention solves the problems inherent in the prior art solutions. For example, the multi-bank memory device of the present invention allows back-to-back read-write operations to the same bank.

[0036] Typically, data in an array must be periodically refreshed. In a read cycle, the data is destroyed and then written back to the memory. A charge in the cell dissipates over time and therefore the cells must be refreshed in order for the data to be retained. During a refresh operation, every row in a DRAM is methodically reviewed and the data is read and written back into the cell (this occurs every few milliseconds). The present invention simulates a process where it appears that a refresh operation is unnecessary by using buffers to bury the refresh operations between read and write cycles. As such, the present invention provides a background refresh operation to take place between the banks, thereby burying the refresh operations as indicated above. This assures that all of the rows in the memory are adequately refreshed. During this background refresh operation, a counter counts the rows, thereby reading the rows. Thus, not all rows are refreshed during one operation as with conventional devices.

[0037] As illustrated in the timing diagram of Figure 3, where Q is data coming out of the memory, and D is write data written into the memory, during the first read operation, the invention accesses memory locations through location address A3 (the memory data must wait because of the inherent latency of DRAM as mentioned above). Then, during the next read cycle, the data comes out in a burst of 4 through address A5.

[0038] Thus, the present invention holds data into output registers, while burst registers hold the data from address A3 to A5. Moreover, the present invention allows the write

data to be presented to the memory two cycles after the write command. Essentially, the present invention lags the read and write cycles. As shown in Figure 3, during the first write cycle, 4 burst bits of data, D2-0, D2-1, D2-2, and D2-3 are held in a buffer. Also, the data coming out of the memory during the first read cycle (at A3), Q0-1, Q0-2, and Q0-3 is read data from the previous read operation.

[0039] Figure 4 illustrates a detailed timing protocol diagram in a multi-bank DRAM implementation with back-to-back read/write operations to the same bank. This illustrates what occurs in the memory (internal RAM operation). For example, there are twelve cycles shown (labeled A through L). The read cycles always occur in the same cycle. Specifically, during the memory operation of Cycle A, a read operation (reading data Q3-0, Q3-1, Q3-2, and Q3-3) for that cycle occurs (i.e., read Cycle A) on a first bank (B1). However, the write operation on the first bank (B1) is lagged, and does not occur until after two write cycles occur on the second bank (B2). Thus, the present invention provides a continuous internal bank swap (B1-B2 flip flop). The CQ represents an echo clock, wherein the memory device outputs Q and CQ data signals and is received by a receiving device, such as a processor or cache, which uses the echo clock to capture the data Q. The refresh cycle operations are sandwiched (buried) in between two bank swaps, as further illustrated in Figure 4. As a result, no refresh-only operations are required.

[0040] The DRAM-specific separate I/O timing protocol described above and illustrated in Figures 3 and 4 illustrate how the high bandwidth can be obtained with a DRAM-based solution. Although the latency is not as aggressive as what a SRAM counterpart offers, the DRAM approach of the present invention provides as high or higher bandwidth (data rate) with a corresponding density increase and reduction in chip size. The DRAM design of the present invention provides the data width necessary for the high frequency bursting of data.

[0041] Figure 5 illustrates a multi-bank memory circuit 1 operable to allow back-to-back read/write memory operations to be performed on a same DRAM bank 3,5, wherein the circuit 1 comprises a plurality of independently addressable DRAM memory banks 3,5; a refresh counter 7 connected to the DRAM memory banks 3,5; a read register 13 connected to the DRAM memory banks 3,5; a write register 15 connected to the DRAM



memory banks 3,5; an address comparator 17 connected to the read and write registers 13, 15, respectively; and a plurality of buffers connected to the DRAM memory banks, wherein the buffers artificially reduces an apparent cycle time of the DRAM memory banks, wherein the refresh counter 7 is operable to send an address 9, 11 to each DRAM memory bank 3,5, wherein the read register 13 is operable to send a read address to each DRAM memory bank 3,5, wherein the write address registers 16 are operable to send a write address to each DRAM memory bank 3,5, and wherein the address comparator 17 is operable to compare the read and write addresses.

[0042] The circuit 1 further comprises a partition array 19, 21 located in each DRAM memory bank 3,5, respectively, and an input clock 27 operatively connected to the read and write registers 13, 15, wherein the input clock 27 is operable to capture the data D, Q from the read and write registers 13, 15. Moreover, the address comparator 17 monitors the write address registers 16 to determine if the write address has been stored in the partition array 19, 21. Also, the refresh counter 7 is operable to track all addresses to be refreshed. Moreover, the read/write memory operations constantly swap between a read and write cycle. As shown, the plurality of DRAM memory banks comprises a first and second DRAM memory bank 3, 5, respectively, wherein the read/write memory operations constantly swap between the first and second DRAM memory banks 3, 5, respectively. This allows the invention to make it appear as though the cycle time is less than the actual cycle time. Furthermore, the read/write memory operations occur in a frequency of  $2n$  times within the same bank 3, 5, again making it appear as though the cycle time is less than the actual cycle time. Finally, as illustrated in Figure 5, an input registering/control unit 25 is provided in the circuit 1 which connects to the input clock 27.

[0043] The invention achieves a number of advantages by grouping read/write access requests. While one group of access requests is being made to a first bank 3, no access requests are made to an alternate second DRAM memory bank 5. Once the group of access requests is completed on the first bank 3, a second group of access requests is applied to the second alternate DRAM memory bank 5. Thus, during a first time period, a first group of access requests is made to a first DRAM memory bank 3 and no access requests are made to a second DRAM memory bank 5. During the next time period, a second group of access requests are made to the second DRAM

memory bank 5 while no access requests are made to the first DRAM memory bank 3.

[0044] Thus, each DRAM memory bank 3, 5 has some "idle time" during which no access requests are being directly made. The invention utilizes this "idle time" to reduce the apparent cycle time of each of the banks 3, 5. Thus, the invention accepts more requests than can be processed during each bank's cycle time and buffers half of the requests. Then, during that bank's idle time (when that bank is not receiving direct accessed requests), the buffered requests can be processed. With two banks, the apparent cycle time can be reduced in half. With four banks operating according to this system, the apparent second time can be reduced to 1/4 the actual cycle time.

[0045] Thus, the invention artificially reduces the apparent cycle time of the different DRAM memory banks 3, 5 through the use of buffers. The invention processes a first request (a read request) and accepts a second request (a write request) before the first bank's cycle time has completed. The invention stores this second request in a buffer and processes it later during its idle time. Then, the invention accepts as a third request (a second read request). At the time of the third request, the bank's cycle time has completed so that it can process the third request. Note that at this time the second request still remains in a buffer. The invention can then take a fourth request (a second write request) before the cycle time has completed, so long as this request is also placed in a buffer.

[0046] A feature of the invention is that while the invention is handling requests for the second DRAM memory bank 5, it has time to process the buffered requests on the first bank 3. This allows the invention to make it appear as though the cycle time is less than the actual cycle time. However, in reality, because the operations are performed in groups, half of the operations are buffered. These buffer operations are executed while requests are made to another bank.

[0047] Using a numerical example, banks 1 and 2 (3, 5) in Figure 5 could have a cycle time of 10 nanoseconds. The invention allows back-to-back read and write requests to occur every 5 nanoseconds. For example, if there were four requests in a group, this would take 20 nanoseconds. Given the cycle time of 10 nanoseconds, two requests of the group could be directly processed and two would be buffered. The first and third requests (read requests) made to bank 1 would be executed while the

second and third requests (write requests) would be buffered. Then, a second group of requests would be made to bank 2. This second group of requests (e.g., 4 requests) also consumes 20 nanoseconds. This 20 nanoseconds is used by bank 1 to process the two buffered write requests. Similarly, during the processing of the second group of requests, bank 2 would buffer two requests which would be processed during bank 2's idle time (during the next group of access requests made to bank 1).

[0048] A method of performing back-to-back read/write memory operations to a same DRAM bank is illustrated in the flow diagram in Figure 6, wherein the method comprises articulating 100 between reading data on a first bank 3 during successive first bank read cycles and writing data to a second bank 5 during successive second bank write cycles, cycling 200 between reading data on the second bank 5 during successive second bank read cycles and writing data to the first bank 3 during successive first bank write cycles, and performing 300 a refresh cycle on the first and second bank 3, 5, respectively, wherein the first bank write cycles lag (are delayed) the first bank read cycles, and wherein the second bank write cycles lag (are delayed) the second bank read cycles. The lag or delay period comprises at least two write cycles.

[0049] As shown in Figure 7, the step of articulating 100 comprises reading 105 data on the first bank 3 during the first bank read cycle, writing 110 data to the second bank 5 during the second bank write cycle, reading 115 additional data on the first bank 3 during the first bank read cycle, and writing 120 additional data to the second bank during the second bank write cycle.

[0050] Furthermore, as illustrated in Figure 8, the step of cycling 200 comprises reading 205 data on the second bank 5 during the second bank read cycle, writing 210 data to the first bank 3 during the first bank write cycle, reading 215 additional data on the second bank 5 during the second bank read cycle, and writing 220 additional data to the first bank 3 during the first bank write cycle.

[0051] The step 300 of performing a refresh cycle occurs on the first bank 3 after the step 120 of writing additional data to the second bank 5. Additionally, the step 300 of performing a refresh cycle is performed on the second bank 5 after the step 220 of writing additional data to the first bank 3.

[0052] Alternatively, as shown in the flow diagram in Figure 9, the present invention provides a method of performing back-to-back read/write memory operations to a same DRAM bank, wherein the method comprises reading 400 data on a first bank 3 during an  $n$ th read cycle, writing 405 data to a second bank 5 during a  $n$ th write cycle, reading 410 data on the first bank 3 during a  $(n + 1)$  read cycle, writing 415 data to the second bank 5 during a  $(n + 1)$  write cycle, refreshing 420 data on the first bank 3, reading 425 data on the second bank 5 during a  $(n + 2)$  read cycle, writing 430 data to the first bank 3 during a  $(n + 2)$  write cycle, reading 435 data on the second bank 5 during a  $(n + 3)$  read cycle, writing 440 data to the first bank 3 during a  $(n + 3)$  write cycle, and refreshing 445 data on the second bank 5.

[0053] In the methods illustrated in Figures 6 through 9, it is shown, according to the present invention, that the read/write memory operations constantly swap between the read and write cycles, and the read/write memory operations also constantly swap between the first and second bank 3,5, respectively. Moreover, the read/write memory operations occur in a frequency of  $2n$  times within the same bank.

[0054] Additionally, according to the present invention, a method of accessing banks of dynamic random access memory (DRAM) memory is provided and shown in Figure 10, wherein the method comprises applying 500 a first group of access requests to a first DRAM memory bank 3 during a first time period, wherein at least one of the access requests is buffered; and applying 505 a second group of access requests to a second DRAM memory bank 5 during a second time period, wherein access requests that are buffered during the first time period are executed during the second time period. The first and second group of access requests comprise read/write memory operations constantly swapping between read and write cycles, wherein the read/write memory operations occur in a frequency of  $2n$  times within each of the first DRAM memory bank 3 and the second DRAM memory bank 5, respectively.

[0055] A representative hardware environment for practicing the present invention is depicted in Figure 11 which illustrates a typical hardware configuration of an information handling/computer system 1000 in accordance with the present invention, having at least one processor or central processing unit (CPU) 1010. The CPUs 1010 are interconnected via system bus 1012 to random access memory (RAM)

1014, read-only memory (ROM) 1016, an input/output (I/O) adapter 1018 for connecting peripheral devices, such as disk units 1011 and tape drives 1013, to bus 1012, user interface adapter 1019 for connecting keyboard 1015, mouse 1017, speaker 1103, microphone 1104, and/or other user interface devices such as a touch screen device (not shown) to bus 1012, communication adapter 1105 for connecting the information handling system to a data processing network, and display adapter 1101 for connecting bus 1012 to display device 1102. A program storage device readable by the disk or tape units is used to load the instructions, which operate the invention, which is loaded onto the computer system 1000.

[0056] There are several benefits of the present invention. For example, the present invention provides a novel DRAM-based separate I/O memory solution for communication applications. Also, the present invention provides a novel DRAM-based solution to the standard QDR communications-SRAM solution. Additionally, the present invention provides data rate and cycle times comparable to conventional QDR SRAMs. Furthermore, the present invention achieves at least four times the density and price/bit of conventional QDR SRAM implementations. Another benefit of the present invention is that it provides a novel banking system whereby back-to-back operations are performed on a given bank.

[0057] While the invention has been described in terms of preferred embodiments, those skilled in the art will recognize that the invention can be practiced with modification within the spirit and scope of the appended claims.